# DOCUMENT RESUME

ED 133 340 TM 005 725

AUTHOR Dziuban, Charles D.; And Others
TITLE An Investigation of Some Distributional Characteristics of the Measure of Sampling Adequacy.
PUB DATE [Apr 76]
NOTE 13p.; Paper presented at the Annual Meeting of the American Educational Research Association (60th, San Francisco, California, April 19-23, 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Analysis of Variance; *Correlation; *Factor Analysis; *Matrices; *Sampling
IDENTIFIERS *Measure of Sampling Adequacy

ABSTRACT

The distributional characteristics of the Kaiser-Rice measure of sampling adequacy (MSA) were investigated with sample correlation matrices from multivariate normal populations where the level of correlation (LC) was systematically varied. Two additional variables were manipulated--sample size (SS) and number of variables (NV). Ten matrices were generated for each LC-SS-NV combination and the overall MSA computed for each, a total of 1,250. Significant effects were found for level of correlation and number of variables as well as for their interaction. Implications for applied factor analysis are discussed. (Author)

ED133340

# AN INVESTIGATION OF SOME DISTRIBUTIONAL CHARACTERISTICS OF THE MEASURE OF SAMPLING ADEQUACY


Charles D. Dziuban

Edwin C. Shirkey

Thomas O. Peeples

Florida Technological University





Presented at the Annual Meeting of the American Educational Research Association

San Francisco

April 19-23, 1976


TM005 725

During 1970 Kaiser announced and later clarified (Kaiser and Rice, 1974) the development of an index which provides information regarding the psychometric adequacy of a data set under consideration for factor analytic methods. His Measure of Sampling Adequacy (MSA) indexes the degree to which the variables in one's investigation comprise an adequate sample from the domain of interest. The overall MSA is defined as:

$$MSA = \frac{\sum\sum_{j\neq k} r^2 jk}{\sum\sum_{j\neq k} r^2 jk + \sum\sum_{j\neq k} q^2 jk}$$

where the $r^2jk$ is a squared off diagonal element of the original correlation matrix and $q^2jk$ is a squared off diagonal of the anti-image correlation matrix, $SR^{-1}S$. The matrix $S^2$ is defined as $\left[\text{diag } R^{-1}\right]^{-1}$. According to present calibrations, MSA lies between zero and one with values in the .80's and .90's signaling data sets which are really appropriate for factor analytic methods. If one should encounter an overall MSA close to .50, it should be clear that the original and anti-image correlations are close to equal ($\sum\sum_{j\neq k} r^2jk = \sum\sum_{j\neq k} q^2jk$). This means that the intercorrelations of the sample anti-image parts of the data are much too large for factoring methods to be considered. (They are uncorrelated in the population).

MSA appears to improve (holding the others constant) as:

(1) The number of variables (NV) increases
(2) The number of subjects (SS) increases
(3) The general level of correlation (LC) increases, and
(4) The effective number of factors decreases

Several studies have been conducted with and of MSA. Dziuban and Shirkey demonstrated (1974A, 1974B) that MSA would signal random variables in a

mixed set (Shaycoft, 1970) and that it would "flag" a correlation matrix based upon random numbers and a sample size of fifty (Armstrong and Soelberg, 1968). They (Shirkey and Dziuban, 1976) conducted a study of the index with sample correlation matrices from populations in which the variables were uncorrelated. Systematic variation of sample size (SS) and number of variables (NV) failed to produce sample-based MSA's which would signal one to proceed with factoring methods. They concluded that there exists little danger that the MSA would erroneously lead one to factor a sample correlation matrix drawn from a population in which the variables were uncorrelated.

Cerny and Kaiser (In Press) contrived matrices which approximated actual sample correlation matrices. They systematically varied the number of variables (p), the number of factors (q), and the root mean square off diagonal correlation (f). They found that the MSA was most sensitive to the number of variables, then the number of factors, and finally the level of correlation. Cerny and Kaiser reported that they with (In Press) Meyer investigated the properties of MSA in the Spearman case as a function of the number of variables and magnitude of correlation. They found that MSA neared .5 when the common off diagonal correlation approached zero and the number of variables approached two.

## Procedure and Results

Sample correlation matrices were created using the Browne (1968) algorythm. The procedure allows one to generate correlation matrices of specified sample size (SS) and number of variables (NV) from a multivariate normal population of a specified correlational level. Ten matrices

(5 x 5 x 5) were generated for each one of the one hundred and twenty-five SS-NV-LC combinations--a total of 1,250.

| | | | | | |
|---|---|---|---|---|---|
| SS = | 50 | 100 | 250 | 500 | 1000 |
| NV = | 5 | 10 | 15 | 20 | 25 |
| LC = | .1 | .3 | .5 | .7 | .9 |

The MSA for each cell of the design was determined. A portion of those means is presented in Table One.

The results of an analysis of variance (Table One) revealed significant ($P < .01$) main effects for number of variables (NV) and level of correlation (LC) as well as their interaction. The strength of association ($\omega^2$) was determined for each significant F ratio. Approximately sixty-eight percent of the MSA variance ($\omega^2 = .677$) was accounted for by the number of variables (NV) and eighteen percent ($\omega^2 = .178$) attributable to level of correlation (LC). The interaction NV x LC accounted for eight percent of the total variance ($\omega^2 = .086$).

---------------------------

Insert Table One About Here

---------------------------

The overall means for the three factors crossed two at a time are presented in Table Two. It may be observed that there was a substantial increase in the MSA for all correlation levels (LC) as the number of variables (NV) increased from five to twenty-five. The nature of that trend may be observed in Graph One. The greatest overall increase was obtained for LC = .1; MSA went from .626 to .956. The increase was not as great for the remaining levels of correlation but the trend continued. At all levels of correlation for NV = 25, the MSA's were substantial (in the high .90's).

Graph Two portrays the change in MSA across level of correlation (LC) for the number of variables (NV). In each case there was an initial increase when LC changed from .1 to .3. For NV = 5 the maximum MSA (.830) was realized at LC = .7. In each of the other cases the highest value was achieved at a level of correlation equal to .5 with slight decreases thereafter. Those decreases, however, would not change the interpretation of MSA levels under the present calibration.

From Tables One and Two it may be noted that sample size had a minimal effect on the index. The overall increase in MSA due to sample size (SS), Table Two, was from .909 to .911. The increase for number of variables was from .777 to .973. Level of correlation resulted in an increase from .839 to .911.

## Discussion

We have found that other things being equal the Kaiser-Rice Measure of Sampling Adequacy tends to be most sensitive to the number of variables (p). Since it is well known that many characteristics of factor analytic solutions are related to "p," our results are not surprising. Increasing the number of variables (NV) from five to twenty-five yields an MSA in the mid to high .90's. With NV = 25, LC = .1 and SS = 50, our $\overline{\text{MSA}}$ was .966. This was quite substantial by present calibration standards.

The MSA is also affected by the general level of correlation of the matrix. The nature of that relationship indicates that increasing LC at the lower levels (.1 to .3) results in a substantial increase in the index. When one has variables correlated approximately at the .5 level, however, the maximum possible MSA will have been achieved.

It is somewhat surprising that the sample size (SS) did not have a stronger effect. It may be that MSA only gets depressed when the sample size (SS) gets down close to the number of variables. This should be verified. In the present study even the smallest sample size (SS = 50) was quite a bit larger than the largest number of variables (NV = 25). It is rare, however, to find practicing factor analysts who work with sample sizes close to fifty, although recently Pruzek (1975) has indicated that small sample factor analysis is indeed a possibility.

It appears from these data that one must expend considerable effort to achieve an MSA as low as .50. Only in the case where the variables are literally unrelated is the MSA likely to approach .50. The lowest cell mean we observed was for LC = .1, NV = 5, SS = 55, MSA = .577. To encounter an MSA in the .50's should be clear warning not to proceed with factor analytic methods. Quite possibly matrices which yield values in the .60's --"not so hot"- should be given serious consideration.

Given the three main effects we have studied, the MSA has performed very predictably with a small surprise regarding the effect of sample size. These results suggest that, if one uses the MSA as the basis of a decision rule for the psychometric quality of his data set, factor analytic practice will be improved.

## References

Armstrong, J. S. and Soelberg, P. On the Interpretation of Factor Analysis. Psychological Bulletin, 1968, 70, 361-64.

Browne, M. A. A Comparison of Factor Analytic Techniques. Psychometrika, 1968, 33, 267-334.

Cerny, B. A. and Kaiser, H. F. A Study of a Measure of Sampling Adequacy for Factor Analytic Matrices. Multivariate Behavioral Research (In Press).

Dziuban, C. D. and Shirkey, E. C. On the Psychometric Assessment of Correlation Matrices. American Education Research Journal, 1974A, 11, 211-16.

Dziuban, C. D. and Shirkey, E. C. When Is a Correlation Matrix Appropriate for Factor Analysis: Some Decision Rules. Psychological Bulletin, 1974B, 6, 358-61.

Kaiser, H. F. A Second Generation Little Jiffy. Psychometrika, 1970, 35, 401-15.

Kaiser, H. F. Little Jiffy Mark IV. Educational and Psychological Measurement, 1974, 34, 111-17.

Meyer, E. P., Kaiser, H. F., and Cerny, B. A. MSA for a Special Spearman Matrix. Psychometrika (In Press).

Pruzek, R. M. Factor Analysis When $n$ is Small: Some Results and Some Optimism. Paper presented at the Annual Meeting of the Psychometric Society, Iowa City, Iowa, April, 1975.

Shaycoft, M. F. The Eigenvalue Myth and the Data Reduction Fallacy. Paper presented at the Annual Meeting of the American Educational Research Association, Minneapolis, Minnesota, March, 1970.

Shirkey, E. C. and Dziuban, C. D. A Note on Some Sampling Characteristics of the Measure of Sampling Adequacy. Multivariate Behavioral Research, 1976, 11, 125-26.

## TABLE ONE

### Results of the Analysis of Variance

| Source | DF | SS | MS | F |
|---|---|---|---|---|
| Subjects | 1249 | 9.3354 | | |
| Number of Variables (NV) | 4 | 6.3806 | 1.5951 | 4077.1[illegible]43 |
| Level of Correlation (LC) | 4 | 1.6632 | .4158 | 1062.7622 |
| Sample Size (SS) | 4 | .0005 | .0001 | .3595 |
| NV x LC | 16 | .8108 | .0506 | 129.5360 |
| NV x SS | 16 | .0088 | .0005 | 1.4148 |
| LC x SS | 16 | .0024 | .0001 | .3878 |
| NV x LC x SS | 64 | .0287 | .0004 | 1.1461 |
| Error | 1125 | .4401 | .C003 | |

*p < .01

$\overline{MSA}$ - NV

| 5 | 10 | 15 | 20 | 25 |
|---|---|---|---|---|
| .777 | .900 | .942 | .962 | .973 |

$\overline{MSA}$ - LC

| .1 | .3 | .5 | .7 | .9 |
|---|---|---|---|---|
| .839 | .927 | .937 | .932 | .919 |

$\overline{MSA}$ - SS

| 50 | 100 | 250 | 500 | 1000 |
|---|---|---|---|---|
| .909 | .911 | .911 | .910 | .910 |

$\overline{MSA}$ - LC x NV

| NV | .1 | .3 | .5 | .7 | .9 |
|---|---|---|---|---|---|
| 5 | .626 | .785 | .829 | .830 | .814 |
| 10 | .799 | .924 | .935 | .929 | .911 |
| 15 | .883 | .963 | .964 | .956 | .943 |
| 20 | .930 | .977 | .976 | .969 | .958 |
| 25 | .956 | .983 | .982 | .976 | .967 |

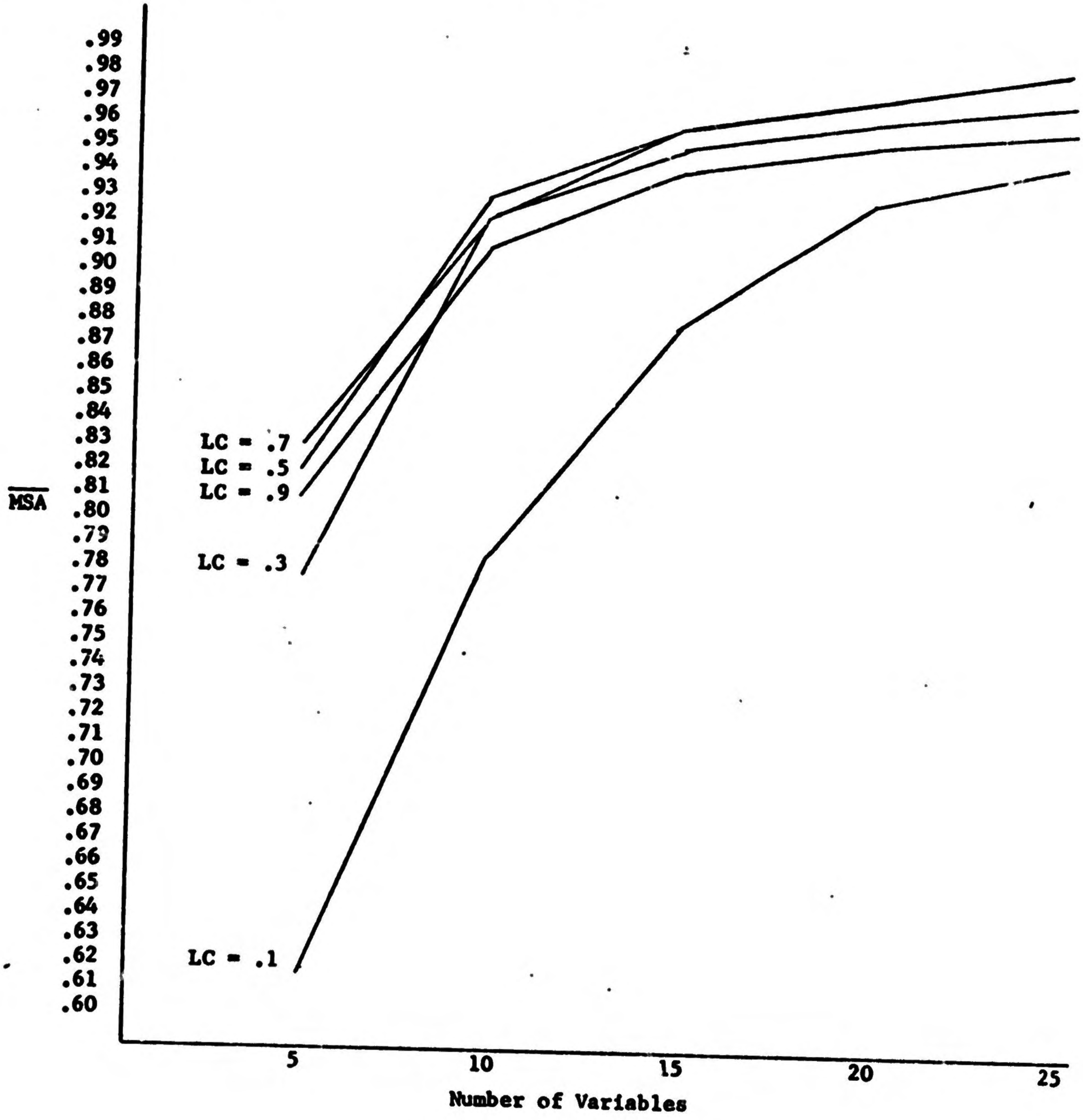
$\overline{MSA}$ - SS x NV

| NV | 50 | 100 | 250 | 500 | 1000 |
|---|---|---|---|---|---|
| 5 | .765 | .779 | .780 | .780 | .779 |
| 10 | .900 | .900 | .900 | .899 | .900 |
| 15 | .943 | .942 | .942 | .941 | .941 |
| 20 | .964 | .963 | .961 | .961 | .960 |
| 25 | .975 | .974 | .972 | .972 | .971 |

$\overline{MSA}$ - LC x SS

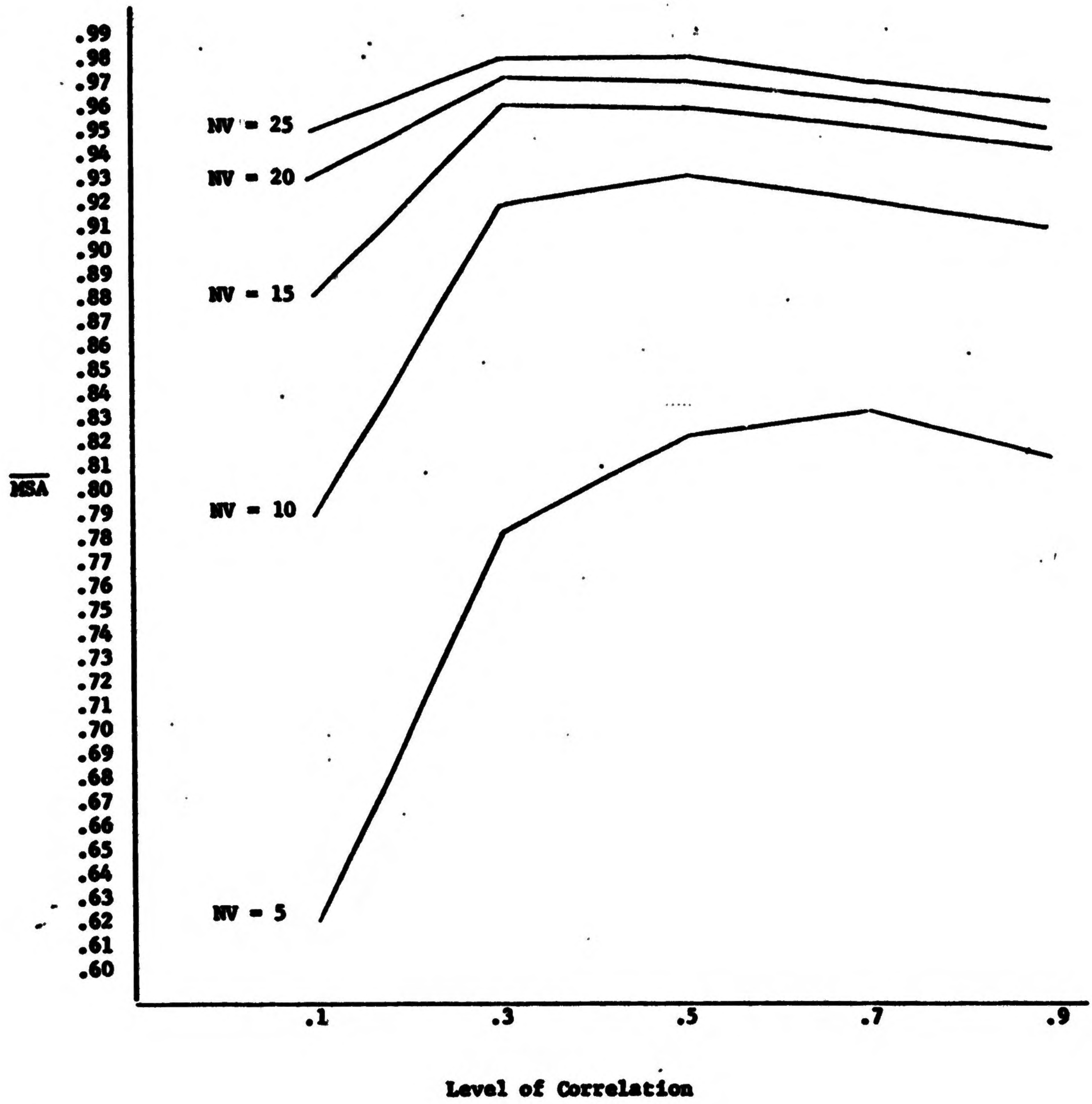
| SS | .1 | .3 | .5 | .7 | .9 |
|---|---|---|---|---|---|
| 50 | .837 | .925 | .936 | .931 | .919 |
| 100 | .844 | .927 | .937 | .930 | .919 |
| 250 | .838 | .928 | .938 | .932 | .919 |
| 500 | .836 | .927 | .937 | .933 | .918 |
| 1000 | .837 | .927 | .937 | .933 | .918 |

Graph I



.99
.98
.97
.96
.95
.94
.93
.92
.91
.90
.89
.88
.87
.86
.85
.84
.83
.82
.81
.80
.79
.78
.77
.76
.75
.74
.73
.72
.71
.70
.69
.68
.67
.66
.65
.64
.63
.62
.61
.60
MSA
LC = .7
LC = .5
LC = .9
LC = .3
LC = .1
5
10
15
20
25
Number of Variables

Graph II



.99
.98
.97
.96
.95
.94
.93
.92
.91
.90
.89
.88
.87
.86
.85
.84
.83
.82
.81
.80
.79
.78
.77
.76
.75
.74
.73
.72
.71
.70
.69
.68
.67
.66
.65
.64
.63
.62
.61
.60
MSA
NV = 25
NV = 20
NV = 15
NV = 10
NV = 5
.1
.3
.5
.7
.9
Level of Correlation